Chapter 8

Discussion and Outlook

8.1 Discussion

In order to realize the sign language recognition system, a software framework to design and test multi-agent systems has been built up. The characteristics of the implemented multi-agent system are autonomous and cooperating units. Principles like divide and conquer, learning from examples and self-control have been applied for object tracking and sign language recognition. Both systems are further divided into smaller subsystems, which are realized as simultaneously running agents. The modular framework allows the recognition system to be easily extensible. New signs are included by connecting their HMM Sensors to a Gesture HMM Integration module which is added to the HMM recognition agent.

The recognition of signs is realized by introducing a modification to the standard HMM architecture. The task of the HMMs is to store feature sequences and their variations. This data is compared with an incoming feature sequence in order to recognize the performed sign. The presented recognition system divides the input features into two types of information. Reliable features show temporal continuity and are more robust under variations between the observed and the learned data, while weaker features are not as robust and will therefore fail more often to be recognized from the observations. Both types of information channels are integrated by using a correlation and rewarding scheme. Another innovation is the competition of the learned signs during the recognition process.

In addition to satisfactory recognition results the autonomy of the system allows to handle the problem of co-articulation. Although the sign rejection experiments did not shown the expected positive rejection rate, the ability of
sign rejection immanent in the design of the recognition system. Therefore it does not need extra modules like threshold or filler HMMs.

Only simple features like the position and hand posture have been applied. The present work does not include a grammar or a high level description. These would be an interesting challenge for future projects and will be discussed in section 8.2.

8.2 Outlook

The presented work can and should be further enhanced by investigating the parameters which control the dynamics and the tracking process as well as the applied thresholds systematically.

Although the system is designed to work online, and thus present the most probable sign for each frame, the recognition processes are too expensive to run in real-time. This problem does not hold if tracking only is demanded. In this case the tracking is running in real-time. In order to speed up the recognition, the modular architecture allows to be simultaneously executed on different computers. First tests using a CORBA interface have successfully been done.

Based on the presented work, sign language research can be continued in the following directions:

Online Learning

In order to have a recognition system, one would like it to be adaptive to sign variations as well as to be capable of learning new signs. The modular design of the present work favors an easy integration of new sign modules. The most challenging task is to build the whole system from scratch by starting with an empty HMM recognition agent. In this case, it would be realistic to start the learning with a defined start and end of the observed sign. Further, under the hypothesis of suitable color segmentation even the hand posture sign lexicons for contour and texture could be build from scratch. Nevertheless, the first step should be to start learning by adding recognized signs to the corresponding sign modules. In a second step, the rejection capability of the system should be enhanced. This allows to find new gestures and then to add them to the HMM recognition agent.

The applied HMM structure allows to expand the distribution probability by just adding a new entry in the histogram for the discrete observations or add a new Gaussian in case of a continuous observation, respectively. Alternatively, the weights of the existing Gaussian mixtures could be adjusted.
to avoid a distribution with too many Gaussians. New states can easily be added to the HMM if the new sign is having more frames than the previously learned ones. However, a problem occurs if the new sign is shorter. In this case, the HMM might not reach the needed confidence threshold which is used to declare the probable occurrence of the sign. As the start of the sign is expected to be given, the computed overall quality might be used to determine the similarity of the observation under the condition that the sign module is active.

**Integration of Non-Manual Information**

The integration of facial expressions is often demanded and important for the full understanding of sign language. However, facial expressions are hard to recognize and should therefore be integrated as a weak feature, giving a reward if the observation match the expected expression or change nothing otherwise. Classification of facial expression is treated in [Tewes et al. (2005)] and could be imported as a new HMM sensor using a discrete feature description.

The consideration of grammar is an important feature for continuous sign language recognition. In the present work, the grammar would not be used to recognize the whole sentence but instead it would predict the appearance of the sign in the context of the previously seen observations. Same as the facial expression, the grammar could be integrated as a weak cue and thus contribute a reward.

**Person Independence**

The most challenging task for sign language recognition systems is the generalization over signer identity. The signer independence capability of the present work could be realized by an enhancement of the hand posture recognition as well as the adaption of the position feature. The trajectories of the hands have to be adjusted to the characteristic behavior of the performing signer.

In order to improve the contour matching, the first task would be to enhance the contour extraction by using enhanced color segmentation. A second approach concerns the process of contour matching which is described in [Horn (2007)]. The author investigates the advantage of integrating contour and texture information as well as the detection of contours. The improvement of the bunch graph concept in order to allow a more generalized hand posture recognition is more difficult. [Triesch and von der Malsburg (2002)] showed that generalization could be achieved if the bunch graph stores
the hand postures of multiple persons. As a second extension, the variations of the landmarks could be learned and stored as a special move which is shown for the facial expressions in Tewes (2006). However, both enhancements of the bunch graph require more human interaction, at least for the initialization.

The adaption of the position information has to solve several problems, because the position of the trajectory in the signing space might differ not only in a linear shift, but as well the whole performance could show nonlinear variations in form and space. The approach to collect more data from different signers to train the HMM is somehow limited. As the variations get to broad the distribution loses its characteristics and becomes less distinguishable from the other signs. Thus, it seems to be necessary to find a transformation that adapts the learned position information to the observed position sequence. This solution would require to solve a global optimization over the learned signs during runtime. In order to limit the number of applied signs, the system profits by the introduction of a grammar and the improved hand posture recognition mentioned above.